Solution structure of domain 5 of a group II intron ribozyme reveals a new RNA motif

Sigel, Roland K O ; Sashital, D G ; Abramovitz, D L ; Palmer, A G ; Butcher, S E ; Pyle, A M

Abstract: Domain 5 (D5) is the central core of group II intron ribozymes. Many base and backbone substituents of this highly conserved hairpin participate in catalysis and are crucial for binding to other intron domains. We report the solution structures of the 34-nucleotide D5 hairpin from the group II intron ai5 gamma in the absence and presence of divalent metal ions. The bulge region of D5 adopts a novel fold, where G26 adopts a syn conformation and flips down into the major groove of helix 1, close to the major groove face of the catalytic AGC triad. The backbone near G26 is kinked, exposing the base plane of the adjacent A-U pair to the solvent and causing bases of the bulge to stack intercalatively. Metal ion titrations reveal strong Mg(2+) binding to a minor groove shelf in the D5 bulge. Another distinct metal ion-binding site is observed along the minor groove side of the catalytic triad, in a manner consistent with metal ion binding in the ribozyme active site.

DOI: https://doi.org/10.1038/nsmb717

The solution structure of domain 5 of a group II intron ribozyme reveals a new RNA motif

Roland K. O. Sigel[1], Dipali G. Sashital[2], Dana L. Abramovitz[3], Samuel E. Butcher[2], Arthur G. Palmer III[3], Anna Marie Pyle[4]

[1] Department of Chemistry, Winterthurerstrasse 190, University of Zürich, CH-8057 Zürich, Switzerland; [2]Department of Biochemistry, University of Wisconsin-Madison, 433 Babcock Drive, Madison, Wisconsin 53706, USA; [3]Department of Biochemistry and Molecular Biophysics, 650 W 168th Street, Columbia University, New York, NY 10032, USA; [4]Department of Molecular Biophysics and Biochemistry and the Howard Hughes Medical Institute, Yale University, 266 Whitney Avenue; New Haven, CT 06520, USA

correspondence should be addressed to A.M.P. anna.pyle@yale.edu, ++1-203-432-5633; ++1-203-432-5316 (FAX) or S.E.B. butcher@nmrfam.wisc.edu, ++1-608-263-3890; ++1-608-262-3453 (FAX)

Abstract

Domain 5 is the central core of group II intron ribozymes. Many base and backbone substituents of this highly conserved hairpin are known to participate in catalysis and to be crucial for binding to other intron domains. The solution structure of the 34 nucleotide domain 5 hairpin from the group II intron ai5γ has been determined by heteronuclear NMR spectroscopy in both the absence and presence of divalent metal ions. The bulge region of D5 adopts a novel fold, in which the bulge guanosine (G26) adopts a *syn*-conformation and flips down into the major groove of helix 1. This conformation brings G26 into close proximity with the major groove face of the catalytic AGC triad. The backbone near G26 is strongly kinked, thus exposing the base plane of the adjacent AU pair to the solvent and causing bases of the bulge to stack intercalatively. Metal ion titrations reveal strong Mg^{2+} binding to a minor groove shelf in the D5 bulge. Another distinct metal ion binding site is observed along the minor groove side of the catalytic triad, in a manner consistent with metal ion binding in the ribozyme active site.

Introduction

Self-splicing of group II introns is essential for the expression of organellar genes in organisms such as plants, yeast, fungi, and bacteria[1,2]. The two-step mechanism employed by these autocatalytic introns for self-excision from pre-mRNA is analogous to that of the spliceosome, which is a dynamic ribonucleoprotein complex that splices nuclear pre-mRNA transcripts in eukaryotes[2,3]. Group II introns possess well-defined secondary structural elements that are organized into six domains (Fig. 1A). Domain 5 (D5) is a highly conserved RNA hairpin consisting of a GAAA tetraloop and two helices that are separated by a bulge region. This hairpin lies at the heart of these ribozymes and is essential for catalysis. Extensive mutagenesis

experiments on D5 have revealed a multitude of atoms and functional groups that are essential for reaction chemistry and also for binding of D5 to other parts of the intron. Functionalities that are directly involved in catalysis[4-7] include the highly conserved AGC triad in helix 1 and a large number of 2'-OH, phosphate oxygens, and base atoms in the bulge region[4,6,8]. The bulge and adjacent nucleotides are very important for function of group II intron ribozymes. Tertiary contact formation through the minor groove of the neighboring GC base pairs in helix 2 brings these region into close contact to D1 and the 5-splice site[8].

Only limited structural information exists on group II intron architecture and domain structure[9,10,11]. The base pairing pattern in the bulge of D5 has been a matter of debate. A C25/G26 bulge has been predicted[12], but recently an alternative secondary structure containing unpaired A24 and C25 and a U9/G26 wobble pair was proposed based on chemical probing[5,13] and sequence covariation analysis[13]. A recently published crystal structure of a modified D5/D6 construct shows a straight, helical D5 hairpin with nucleotides A24 and C25 flipped out and U9/G26 in a wobble pair[9]. However, the bulge region makes lattice contacts to sister molecules, which may cause these nucleotides to adopt the observed conformation.

Recently, high affinity metal ion binding sites in a group II intron were identified by lanthanide(III) mediated hydrolysis experiments[14]. A major metal ion binding site was revealed in the bulge of D5 and adjacent nucleotides. The Ln^{3+} cleavage experiments also showed that the metal ion binding observed in the bulge of isolated D5 molecules persists in the bound form, in context of domains 1, 3, and 5, suggesting that D5 alone adopts the active structure[14]. However, due to conformational constraints on the Ln^{3+} cleavage reaction, additional strong metal ion binding sites could remain undetected through this assay. In a study that evaluated factors that contribute to group II intron catalysis, a metal ion was detected in coordination with the pro-S_p oxygen of A2[7], and found to play a role in D5 binding.

To shed more light on the structural and metal ion binding properties of D5, we have solved the solution structure of this domain in the presence of Mg^{2+}. Interestingly, the bulge region adopts a hitherto unknown motif, exposing the hydrophobic surface of an AU base pair to solvent. Furthermore, data from Mg^{2+} titrations that were monitored by NMR spectroscopy are consistent with localization of a metal ion in the exposed minor groove surface of the bulge and along the minor groove backbone near the catalytic triad.

Spectral Features and General Characteristics of D5

The RNA construct we chose for NMR studies corresponds to the wild type sequence of domain 5 (D5: nucleotides 815-848) from group II intron ai5γ, which is located in the *cox1* gene of *S. cerevisiae* (Fig. 1A). We extended the hairpin by an additional GC base pair (numbered G0 & C35) to stabilize the helix terminus and to improve the yield from *in vitro* transcription. Comparison of 2D NOE data of the extended and the wild type (wt) D5 demonstrated that this minor modification has no influence on the structure of the nearby catalytic AGC triad, nor does it affect catalytic activity (data not shown). This additional base pair was not included in the structure calculation.

NMR spectra directly reveal overall structural features of D5. 1D ^{1}H spectra acquired in H_2O at temperatures between 273 K and 318 K and with varying metal ion concentrations (0-200 mM KCl and/or 0-20 mM $MgCl_2$) consistently show resonances of slowly exchanging imino protons belonging to ten Watson-Crick base pairs, two GU wobble pairs and an additional imino of G15 in the GAAA tetraloop at about 10.5 ppm. All observable imino proton resonances were assigned by 2D $^{1}H,^{1}H$ NOESY and $^{1}H,^{15}N$ HSQC spectra. Imino protons of U9, G26, and U27, which are located in the bulge region, were not observed under any conditions tested due to fast exchange of the protons with the solvent. This suggests that the bulge adopts an exposed

structure (see also below). The imino proton of the terminal G0 was also not observed. Base pairing for all GC and AU base pairs in the two helical regions was confirmed by J_{NN} HNN-COSY experiments[15,16] in D_2O and H_2O (data not shown). Despite the absence of an observable imino proton resonance for U27, crosspeaks between H2 of A8 and N3 of U27 in J_{NN} HNN-COSY experiments verified this Watson-Crick base pair adjacent to the bulge. However, U9, A24 and C25 are clearly positioned in the helix, as confirmed by characteristic sequential and stacking NOEs as well as a distinct interstrand NOE between A24H2 and G10H1'. NOE resonances for the bulge and adjacent nucleotides are only detectable at 500 and 600 MHz. At higher field (750 MHz), these resonances are broadened, indicating internal motion within the bulge region (data not shown). Similar such field-dependent relaxation has been observed in a dynamic RNA pseudoknot structure[17].

Sequential NOEs link nucleotides from the 5'-G0 to G26 in the bulge and from U27 to the terminal C35 at the 3'-end of the hairpin (Fig. 1B). The missing correlation for G26H8-U27H1' interrupts the sequential walk at the 3'-end of the bulge. Instead, H8 of G26 exhibits an internucleotide NOE to H1' of A28 indicating an unusual geometry at the bulge. An unusually strong intranucloetide H8-H1' NOE resonance for G26 suggests a *syn*-conformation of the purine moiety and provides additional evidence for deviations from an A-form helix. The *syn*-geometry was confirmed by a short mixing time $^1H,^1H$ NOESY and the absence of intranucleotide NOE resonances from G26H8 to H2' and the other sugar protons at longer mixing times.

The catalytic AGC triad near the hairpin terminal has been proposed to be a region of enhanced flexibility[9]. In solution, no evidence for such flexibility is observed. A stable base pairing pattern was observed for all three nucleotides under all conditions tested. This agrees with chemical modification experiments showing that G3 is buried within the helix in context of the whole intron[5].

Based on NMR data described above, a regular A-form geometry for helices 1 (G1/C34 to A8/U27) and 2 (G10/C23 to U14/G19) was assumed for structure calculation. The D5 hairpin is closed on one end by a GAAA tetraloop. Chemical shifts and NOEs for the GAAA nucleotides indicate a very stable structure with the second nucleotide (A16) tightly stacked.

Structure determination of domain 5

The structure determination of D5 is based on 678 conformationally restrictive NOE distance restraints and 24 residual dipolar coupling (RDC) constraints (Table 1). D5 adopts a stable hairpin structure with two well defined helical regions that are separated by the bulge nucleotides (Fig. 2A). The same overall conformation of the hairpin was obtained with or without the RDC restraints. However, the RDC restraints help to define the angle between the two helical axes and were therefore included in the final calculation. A superposition of the final ten lowest-energy conformers over all heavy atoms is shown in Supplementary Fig. 1. The two helices of D5 adopt a standard A-form geometry and, together with the GAAA tetraloop, do not differ from the recently published crystal structure[9] or previously published structures of the GAAA tetraloop[18,19]. The r.m.s. deviations over all heavy atoms of the ten lowest-energy structures is 4.15 ± 1.38 Å (see also Table 1). However, independent superpositions of the bulge region, helix 1 or helix 2 result in much lower r.m.s. deviations (Table 1 and Supplementary Fig. 1). Thus, the rather high total r.m.s. deviation is due to the observed flexibility in the bulge region, which acts like an elbow between the two helices.

The bulge adopts a hitherto unknown structure

In solution, the bulge region differs both from previously proposed[5,12,13] and crystallographically-determined [9] structure (Fig. 3). Stacking interactions appear to govern the bulge geometry,

because distinct hydrogen bonds between the bulge nucleotides are not observed in solution. The majority of distinguishing features were observed in all ten of the lowest-energy structures (Fig. 3A). Helix 1 is closed by the A8/U27 base pair. U9 remains stacked onto the imidazole ring of A8 and to a lesser extent onto G10. On the opposite strand, A24 and C25 are stacked on the terminal G10-C23 base pair, thus elongating helix 2 on the 3' side. A24 is intercalated between U9 and G10 causing a tilt between the nucleobase planes of U9 and G10 and bringing the backbones of the two strands closer together. The adjacent *syn*-G26 is flipped down into the major groove of helix 1 and lies directly on top of the catalytic AGC triad. As a consequence, helix 1 and 2 are contorted towards each other and a considerable kink similar to an "S" is observed at nucleotides A24-U27 (Fig. 3B). Due to this distortion of the backbone, the A8/U27 base pair in helix 1 is exposed to the solvent. This conformation explains the absence of an observable imino proton resonance for U27. Such an extensive exposure of hydrophobic base surfaces to solution within a well-ordered RNA structure is very unusual and has, to the best of our knowledge, only been observed to a lesser degree in the U-turn platform within the specificity domain of RNAse P[20].

All spectral data concur with the distinct bulge conformation that results from the structure calculation. A detailed analysis of distances between non-exchangeable protons of the bulge nucleotides revealed that all expected NOE crosspeaks were found in the NOE spectra. Although the lowest energy structures share most of the major features, there are some differences among them. Superimposition of the bulge region reveals the bases are loosely stacked (Fig. 3B). We observe that alternative hydrogen bonding patterns within the bulge are possible, though in none of the structures is a U9/A24 base pair observed as it has been proposed originally[12]. Using a distance criteria of <3.5 Å between donor and acceptor heteroatoms, half of the lowest energy structures show U9 and C25 connected by at least one hydrogen bond (U9N3H-C25N3, U9N3H-C25O2, and/or U9O4-C25N4H_2). Interestingly, in three structures C25

not only hydrogen bonds to U9 but is also close enough to the Hoogsteen face of A8 to form hydrogen bonds and thus form a putative UAC base triple in a trans Watson-Crick/Hoogsteen fashion. Attempts to corroborate any of these putative hydrogen bonding patterns by scalar coupling experiments[15] were unsuccessful. An ensemble of structures with different base pairing and hydrogen bonding patterns is likely to exist in solution. Intermediate time scale exchange between conformations could explain the field dependent relaxation phenomenon observed in this region of D5.

The structure of the bulge is unusual in two respects. First, unpaired nucleotides of a RNA helix are typically flipped out into solution or alternatively embedded into the minor groove. Instead, in D5 three of the four unpaired nucleobases are stacked within the helix and *syn*-G26 is flipped down into the major groove. Second, exposure of the hydrophobic surface of a base pair within a RNA hairpin is surprising on first sight. However, most of the nucleobase hydrophobic surfaces of D5 are buried by accommodating many of the nucleotides within the structure. A reason for the formation of such a structure might be that the two nucleotide bulge of D5 is too big to be flipped out and to allow the stable formation of an extended hairpin in solution.

The structure of D5 has immediate biological implications. G26 is not only strongly conserved[12,21,22], but also exhibits a strong interference effect on catalysis when switched to an inosine moiety[23]. Because G26 has a *syn* conformation and is flipped down into the major groove of helix 1, the hydrogen-bonding groups of G26 are in the vicinity of chemically essential elements of the catalytic AGC triad. The Watson-Crick face of G26 and the 2'-OH groups of G26 and A24 have been shown to be important for catalysis[4,5] and strong phosphorothioate effects have been observed for A24 and C25[6,7,23]. The tandem GC base pairs above the bulge in helix 2 compose the important λ-λ' tertiary contact within the intron[8]. Thus, the minor groove pocket and shelf created by the bulge conformation provides an ideal hydrophobic platform for stacking interactions with incoming nucleotides from other intronic elements, similar to an AA platform of

a tetraloop receptor[24]. The observed flexibility in the bulge region might either enable D5 to dock effectively into D1 or facilitate structural changes during the catalytic pathway.

Metal ion binding to D5

Because D5 is packed tightly in the catalytic center of group II introns[11], D5 may bind metal ions required for catalysis or to overcome electrostatic repulsion within the core of the folded intron. We performed Mg^{2+} titration experiments to further characterize the metal ion binding sites in the D5 bulge region and adjacent to the AGC triad, and to investigate the existence of other Mg^{2+} binding sites. Titrations can be performed by monitoring changes in chemical shifts of imino protons upon addition of Mg^{2+}. However, to obtain a particularly detailed picture of Mg^{2+} binding to all residues in D5, we also recorded $^{1}H,^{1}H$-NOESY spectra in D_2O. The concentration of Mg^{2+} was increased stepwise from 0 to 7.5 mM (corresponding to 15 molar equivalents of D5 RNA). Chemical shift changes were followed for all aromatic, H1', H2', H3', as well as for a significant number of H4' and H5'/H5" protons for every nucleotide in D5. Most of these non-exchangeable protons are much closer to potential metal ion coordination sites (such as N7 of purines, and the phosphate oxygens) than imino protons. With this approach, metal ion binding in the bulge could be evaluated despite the lack of observable imino proton resonances in this region. All NOE correlations could be observed up to a Mg^{2+} concentrations of 7.5 mM.

Analysis of chemical shift changes for aromatic protons as well as H1' and H2' indicates that Mg^{2+} affects the bulge nucleotides more strongly than any other region of D5 (Fig. 4). Major shift changes occur primarily on minor groove constituents of the bulge, e.g. the H1' sugar protons of U9 and C25, the H2' of G10, C23, A24, C25 and U27, as well as the aromatic H2 protons of A8 and A24 (Fig. 4). The majority of potential binding sites for hard ions like Mg^{2+} are located in the minor groove, which creates a shallow bowl immediately above the A8-U27

platform. This space includes the strongly shifted H5 of U27, and the hard carbonyl oxygens U9O2, C25O2 and U27O4, which are close enough to be bridged by a single metal ion. Resonances of protons in the major groove of the bulge (H8 of G10 and A24, as well as H5 and H6 of C25) also are affected by the Mg^{2+} binding and therefore metal ion interactions in the major groove cannot be ruled out. However, chemical shifts can be affected either by direct metal ion binding or by the structural changes that binding induces, i.e. reduced or increased stacking interactions. The most likely explanation for chemical shift changes in major groove protons (including H5 and H6 of C25, which are distant from any potential metal binding site) is a slight disturbance in local geometry. Although the chemical shift data strongly implicate the minor groove bulge region as a metal ion binding site, the lack of a uniform pattern of the chemical shift data makes it difficult to interpret the titration data in terms of a precise coordination site for metal ions. Interestingly, the adjacent G26 is almost unaffected by the addition of Mg^{2+}. This is consistent with the burial of G26 base in the major groove where it makes Van der Waals contact with C29, which is also unaffected by Mg^{2+}. The possibility of multiple metal ion binding sites in the bulge region must also be taken into account. To further investigate these issues, we calculated binding constants for Mg^{2+} from our titration data. All titration data could be fit to a 1:1 binding isotherm, suggesting that indeed only one Mg^{2+} ion binds to the bulge region. Evaluation of the chemical shift data revealed the strongest binding site at C25 (K_D = 3.0 ± 0.3 mM; which corresponds to the mean value of the individual K_D^* values for the aromatic and sugar protons, see also Supplementary Fig. 2). Other binding constants for this region vary between 3.5 and 6.0 mM.

To further investigate metal ion coordination in the bulge region, we used the non-linear Poisson-Boltzmann equation to calculate the electrostatic surface potential of D5[25]. A small region of intensely negative electrostatic potential (–60 kT/e) is located in the minor groove bowl of the bulge that includes the O2 oxygens of U9 and C25, as well as N3 of G10 and A24 (Fig.

2B) This large deviation in electrostastic potential, together with the significant chemical shift changes in the same minor groove region of the bulge are strong indicators of Mg^{2+} binding in this pocket.

Aside from the bulge region, smaller chemical shift changes are also observed in the tetraloop region and at C29, which is part of tandem GC base pairs in helix 1 (Fig. 4). A fit of the chemical shift data to a 1:1 binding isotherm revealed that strong Mg^{2+} binding occurs near C29 (K_D = 1.2 ± 0.7 mM). Indeed, the major groove of tandem GC base pairs is known to bind metal ions[26]. This corresponds well to the observed high negative electrostatic potential in these regions (Fig. 2B).

The catalytic triad (A2, G3 & C4) of D5 is another region of special interest. As already described above, the NOE pattern and base pairing in the triad do not change upon addition of up to 20 mM Mg^{2+}. No significant chemical shift changes of nucleobase or sugar protons were observed in this region (Fig. 4). To further investigate metal ion binding at the AGC triad, we performed paramagnetic line broadening experiments with Mn^{2+}. Indeed, at low micromolar concentrations of $MnCl_2$, a small set of resonances near the 5' end of the D5 hairpin were broadened. No effects are seen at other parts of D5. In 45 μM Mn^{2+}, all NOEs for G0 and G1H8 disappear and the H1' of G1 and A2 become very broad, whereas resonances on the opposite strand are basically unaffected (see also Supplementary Fig. 3). At concentrations above 45 μM Mn^{2+}, G3H8 of the catalytic triad also becomes substantially broadened. Taken together these data suggest that Mn^{2+} (and therefore probably Mg^{2+}) does not bind directly to the major groove of the catalytic triad, i.e. where catalysis has been suggested to take place[5]. The fact that the A2H1' is more affected by the Mn^{2+} ion than the aromatic protons of the AGC triad suggests that metal ion binding takes place along the adjacent minor groove backbone.

These effects of metal ion binding to the isolated D5 molecule are particularly significant because metal ion binding has been observed in the same region of D5 when it is part of an activeribozyme core. In functional group mutagenesis experiments on catalysis by a group II intron, the pro-S_p oxygen of A2 was shown to bind a metal ion that is important for binding of D5 to D1[7]. Similar correlations between metal ion binding to isolated D5 and D5 bound in the functional state was also observed during Tb^{3+} cleavage experiments[14]. Taken together, the data on metal ion interactions in D5 establish that metal ions play an important role in the structure and long-range tertiary interactions of D5 by binding in two functionally important locations along the minor groove. It is notable that metal ions have not been detected in the major groove of the AGC triad , nor at G26 that docks nearby. The major groove in this region contains the group II intron atoms that have been most directly implicated in reaction chemistry. This suggests either that catalytic metal ions do not enter the AGC major groove until D5 is fully docked within the ribozyme core, or that D5 participates in a chemical mechanism that is, at least in part, metal ion independent.

Conclusion

The solution structure of D5 differs substantially from the recently published crystal structure of D5/D6[9]. In solution, where D5 is not influenced by lattice contacts, all bulge nucleotides are positioned within the helical structure, except for G26 which extrudes into the major groove next to the AGC catalytic triad. The bulge is a region of high flexibility that acts like a hinge between the helix arms possibly enabling a rearrangement in the catalytic core between the first and the second step of splicing. Nevertheless, the conformation of the bulge nucleotides in solution forms a very interesting and, to best of our knowledge, a hitherto unknown RNA motif. A WORM search[27] and further extensive search in the NDB database did not reveal a similar RNA motif

with a nucleotide buried within the major groove. However, several lines of data make it worthwhile to consider our D5 structure also to be present in such a conformation within the active intron. First, the buried G26 is highly conserved[12], and functionally important[23]. Our solution structure now positions this purine close to the almost invariant catalytic AGC triad, whose major groove functionalities have been identified to participate in catalysis[5]. Second, Tb^{3+} hydrolysis experiments showed that metal ion binding in the tetraloop as well as at the bulge and adjacent nucleotides persists, when D5 is examined in isolation or in context of the whole intron. Together with the exposed hydrophobic surface of the adjacent A8/U27 bp, and the λ-λ' hydrogen bonding network, a metal ion in the minor groove of the bulge might enable D5 to bind to catalytic regions in D1 and the 5'-splice site.

Materials and Methods

NMR sample preparation. D5 RNA with an additional GC base pair at the open end of the hairpin was transcribed *in vitro* using T7 RNA polymerase (Stratagene) and synthetic, double stranded DNA oligonucleotides as templates (Protein Core Facility, HHMI, Columbia University). RNA was purified by denaturing 17% PAGE, identified by UV shadowing and excised from the gel. The RNA was then recovered by electroelution (Schleicher & Schuell, Inc.), ethanol precipitated and exchanged into NMR buffer by at least four spins through Centricons YM-3 (Amicon Bioseparation, Millipore) at 7000 g. pH was then adjusted with 0.1 M KOH (if not stated otherwise, buffer (pH 6.5) contained 100 mM KCl, and 10 μM Edta). RNA concentrations were between 0.5 and 0.7 mM. All samples were lyophilized and resuspended in either 90% H_2O/10% D_2O or 99.999% D_2O (Aldrich Inc.). ^{13}C,^{15}N-labeled NTPs were purchased (Isotec, Martek Biosciences Corp., and Silantes GmbH.) and ^{13}C,^{15}N-labeled D5 was then prepared accordingly.

NMR spectroscopy. NMR spectra were recorded on a Bruker Avance DMX 600 MHz (Columbia University, NY), and on Bruker DMX spectrometers (750 & 500 MHz) at the National Magnetic Resonance Facility at Madison, WI (NMRFAM). The 500 MHz machine was equipped with a HCN triple-resonance z-axis pulsed-field-gradient cryoprobe™. Other spectrometers were equipped with HCN triple-resonance triple-axis pulsed-field gradient probes.

Exchangeable resonances were assigned by reference to 2D ^{1}H,^{1}H-NOESY (250 ms mixing time, 278 K), and ^{1}H-^{15}N HSQC (303 K) spectra of the RNA in 90% H_2O/10% D_2O. Non-exchangeable resonances were assigned by reference to 2D ^{1}H,^{1}H-NOESY spectra (60, 120, and 250 ms mixing times), ^{1}H,^{1}H-TOCSY (44 ms mixing times), and ^{1}H-^{13}C HSQC and 3D ^{1}H-^{13}C-^{1}H HCCH TOCSY, ^{1}H-^{13}C-^{1}H HCCH COSY and ^{1}H-^{13}C-^{1}H NOESY-HMQC (250 ms mixing time) spectra of the RNA in D_2O at 298 K. 2D NOESY spectra also were recorded at 291, 303, and 308 K. Water suppression for samples in 90% H_2O/10% D_2O was achieved with a 1-1 spin-echo pulse sequence. For experiments in D_2O, the residual HDO resonance was suppressed with a low power presaturation pulse. Quantitative J-correlation $^{2}J_{HN}$ HNN-COSY[15] and dual-detected J_{NN} HNN-COSY[16] experiments were used to confirm base pairing in D5.

Partial alignement of RNA for residual dipolar coupling (RDC) measurements was achieved by adding 17 mg/mL Pf1 filamentous bacteriophage (ASLA Ltd., Riga, Latvia) to the uniformly ^{13}C- and ^{15}N-labeled sample[28,29]. ^{1}H-^{13}C-RDCs were measured in the carbon (F1) and proton (F2) dimension of 2D ^{1}H-^{13}C CT-HSQC spectra at 303 K for both isotropic and oriented samples. Uncertainty in RDC measurements was estimated to be ±2.0 Hz.

NMR data were processed with XWINNMR (Bruker) and analyzed with the NMR assignment program Sparky[30]. NOE peak volumes were integrated with the Gaussian peak fitting function in Sparky[30].

Structure Calculation. NOE distances were estimated from the integrated peak volumes obtained from a NOESY spectrum in D_2O at 303 K with a 250 ms mixing time, which was determined to be within the linear range of the NOE build-up curve. Distances were calibrated by setting the average integrated volume of the pyrimidine H5-H6 NOEs to 2.4 Å, using the r^{-6} relationship and the CALIBA macro in DYANA[31]. NOEs were then grouped into four categories, corresponding to strong (1.8 – 3.0 Å), medium (1.8 – 4.5 Å), weak (3.0 – 6.0 Å), and very weak (4.0 – 7.0 Å). NOEs from exchangeable protons obtained in 90% H_2O/10% D_2O were qualitatively assigned as strong, medium, weak or very weak.

The torsion angles α and ζ were loosely restrained to exclude the *trans* range (0 ± 120º). Based on the integrated NOE volume of intranucleotide H1' to aromatic NOE for a 60 ms NOESY spectrum, the torsion angle χ was restrained to 70 ± 30º for G26 (*syn*) and for all other nucleotides to –160 ± 10º (*anti*). TOCSY experiments with a 45 ms mixing time were used to analyze sugar pucker conformations. Nucleotides with strong H1'-H2' and H1'-H3' crosspeaks were restrained to S-type range (155 ± 10º) (A16 and C25). Nucleotides with intermediate crosspeak intensities (G15, A17, A18) were left unrestrained. Nucleotides with absent H1'-H2' crosspeaks were restrained to N-type range (85 ± 10º).

Residual dipolar coupling constants (RDCs) were measured using Sparky[30] by determining the difference between ^{1}H-^{13}C coupling for isotropic and partially aligned samples. PALES[32] was used to estimate values for the axial and rhombic components (D_a and R) of the alignment tensor from low energy structures calculated in the absence of RDCs, yielding D_a = -30 and R = 0.34. CNS structure calculations with a grid search for optimal D_a and R values revealed that the best agreement between structure and RDCs, as evident from low energies and no violation, is obtained with D_a = -28.5 and R = 0.24.

CNS 1.1[33] was used to calculate structures using NOE distance, dihedral, and RDC constraints. CNS was recompiled with an improved version of the harmonic potential for RDCs, which corrects the susceptibility anisotropy refinement protocol[34]. First, an extended structure (completely unfolded) structure was generated, from which 100 starting structures were calculated from random initial velocities. The 100 starting structures were subjected to 60 ps of restrained molecular dynamics in torsion angle space, using 15 fs time steps, followed by 90 ps of slow cooling. Finally, 30 ps (5 fs time steps) of restrained molecular dynamics in Cartesian coordinate space was performed. Planarity for all base pairs detected using exchangeable 2D NOESY and trans-hydrogen bond experiments was enforced during the calculations, and hydrogen bonds were maintained by distance restraints for Watson-Crick base pairs as well as one GA and two GU pairs. Structures were also calculated without RDCs for comparison. After Cartesian space refinement, the structures were evaluated for convergence. Acceptance criteria of converged structures were low overall energies and no significant NOE (>0.5 Å) or dihedral (>5°) violations.

The 10 lowest energy refined structures were viewed and analyzed using MOLMOL[35]. Four additional structures of low overall energies were found that satisfied all the RDC and NOE restraints, but these could be discarded due to near anti-parallel alignment between the two helices, owing to the fact that the RDC restraints can be satisfied by both A-form geometry or anti-parallel helical orientations that are rotated180° about the order tensor frame S_{xx} and S_{yy}[36].

Structure Analysis. The electrostatic surface potentials were calculated using QNIFFT [25] and visualized with GRASP [37]. Conformational analyses of the RNA structure, comparison with the x-ray structure of D5, and motif searches were performed with the AMIGOS[38] and WORM[27] algorithms.

Metal Ion Titrations and Calculation of Stability Constants. Mg^{2+} binding to D5 was evaluated by observing the changes of chemical shifts of aromatic and sugar protons upon increasing metal ion concentration. Small aliquots of $MgCl_2$ (100 mM and 1 M stock solution in D_2O) were successively added to a solution of D5 RNA to a final Mg^{2+} concentration of 7.5 mM. Changes of chemical shifts were monitored by recording $^1H,^1H$-NOESY spectra (D_2O) at each titration point. Chemical shift changes were plotted against Mg^{2+} concentrations and fit to a single binding isotherm by Levenberg-Marquardt non-linear least-squares regression. Chemical shift changes of imino protons were monitored by 1D spectra in 90% H_2O/10% D_2O (concentration range of Mg^{2+}; 0 – 18 mM).

Acknowledgements

We thank Laura Nikstad for helping with the structure determination, and Phil Pang, Carlos Duarte & Leven Wadley for many helpful discussions. NMR studies carried out at the National Magnetic Resonance Facility at Madison were supported from the NIH Biomedical Technology Program and additional equipment funding from the University of Wisconsin, NSF Academic Infrastructure Program, NIH Shared Instrumentation Program, NSF Biological Instrumentation Program, and the U.S. Department of Agriculture. We are very grateful for financial support from the Swiss Academy of Natural Sciences and the Swiss National Science Foundation (Grant for Young Scientists for R.K.O.S.), the NSF (A.G.P.) and the NIH (S.E.B. and A.M.P.). A.M.P. is an Investigator of the HHMI.

References

1. Bonen, L. & Vogel, J. The ins and outs of group II introns. *Trends Genet.* **17**, 322-331 (2001).
2. Michel, F. & Ferat, J.L. Structure and activities of group II introns. *Annu. Rev. Biochem.* **64**, 435-461 (1995).
3. Sontheimer, E.J., Gordon, P.M. & Piccirilli, J.A. Metal ion catalysis during group II intron self-splicing: parallels with the spliceosome. *Genes Dev.* **13**, 1729-1741 (1999).
4. Abramovitz, D.L., Friedman, R.A. & Pyle, A.M. Catalytic role of 2'-hydroxyl groups within a group II intron active site. *Science* **271**, 1410-1413 (1996).
5. Konforti, B.B. et al. Ribozyme catalysis from the major groove of group II intron domain 5. *Mol. Cell* **1**, 433-441 (1998).
6. Chanfreau, G. & Jacquier, A. Catalytic site components common to both splicing steps of a group II intron. *Science* **266**, 1383-1387 (1994).
7. Gordon, P.M. & Piccirilli, J.A. Metal ion coordination by the AGC triad in domain 5 contributes to group II intron catalysis. *Nat. Struct. Biol.* **8**, 893-898 (2001).
8. Boudvillain, M., deLencastre, A. & Pyle, A.M. A tertiary interaction that links active-site domains to the 5' splice site of a group II intron. *Nature* **406**, 315-318 (2000).
9. Zhang, L. & Doudna, J.A. Structural insights into group II intron catalysis and branch-site selection. *Science* **295**, 2084-2088 (2002).
10. Costa, M., Michel, F. & Westhof, E. A three dimensional perspective on exon binding by a group II self-splicing intron. *EMBO J.* **19**, 5007-5018 (2000).
11. Swisher, J., Duarte, C.M., Su, L.J. & Pyle, A.M. Visualizing the solvent-inaccessible core of a group II intron ribozyme. *EMBO J.* **20**, 2051-2061 (2001).

12. Michel, F., Umesono, K. & Ozeki, H. Comparative and functional anatomy of group II catalytic introns - a review. *Gene* **82**, 5-30 (1989).

13. Costa, M., Christian, E.L. & Michel, F. Differential chemical probing of a group II self-splicing intron identifies bases involved in tertiary interactions and supports an alternative secondary structure model of domain V. *RNA* **4**, 1055-1068 (1998).

14. Sigel, R.K.O., Vaidya, A. & Pyle, A.M. Metal ion binding sites in a group II intron core. *Nat. Struct. Biol.* **7**, 1111-1116 (2000).

15. Hennig, M. & Williamson, J.R. Detection of N-H-N hydrogen bonding in RNA via scalar couplings in the absence of observable imino proton resonances. *Nucleic Acids Res.* **28**, 1585-1593 (2000).

16. Luy, B. & Marino, J.P. Direct evidence for watson-crick base pairs in a dynamic region of RNA structure. *J. Am. Chem. Soc.* **122**, 8095-8096 (2000).

17. Kolk, M.H. et al. NMR structure of a classical pseudoknot: interplay of single- and double-stranded RNA. *Science* **280**, 434-438 (1998).

18. Heus, H.A. & Pardi, A. Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science* **253**, 191-194 (1991).

19. Rüdisser, S. & Tinoco, I., Jr. Solution structure of cobalt(III)hexammine complexed to the GAAA tetraloop, and metal-ion binding to G·A mismatches. *J. Mol. Biol.* **295**, 1211-1223 (2000).

20. Krasilnikov, A.S., Yang, X., Pan, T. & Mondragon, A. Crystal structure of the specificity domain of ribonuclease P. *Nature* **421**, 760-764 (2003).

21. Ferat, J.L. & Michel, F. Group II self-splicing introns in bacteria. *Nature* **364**, 358-361 (1993).

22. Chu, V.T., Adamidi, C., Liu, Q., Perlman, P.S. & Pyle, A.M. Control of branch-site choice by a group II intron. *EMBO J.* **20**, 6866-6876 (2001).

23. Boudvillain, M. & Pyle, A.M. Defining functional groups, core structural features and inter-domain tertiary contacts essential for group II intron self-splicing: a NAIM analysis. *EMBO J.* **17**, 7091-7104 (1998).

24. Basu, S. et al. A specific monovalent metal ion integral to the AA platform of the RNA tetraloop receptor. *Nat. Struct. Biol.* **5**, 986-992 (1998).

25. Chin, K., Sharp, K.A., Honig, B. & Pyle, A.M. Calculating the electrostatic properties of RNA provides new insights into molecular interactions and function. *Nat. Struct. Biol.* **6**, 1055-1061 (1999).

26. Schmitz, M. & Tinoco Jr., I. Solution structure and metal-ion binding of the P4 element from bacterial RNase P RNA. *RNA* **6**, 1212-1225 (2000).

27. Duarte, C.M., Wadley, L. & Pyle, A.M. RNA structure comparison, motif search and discovery using a reduced representation of RNA conformational space. *Nucl. Acid. Res.*, in press.

28. Zhou, H.J., Vermeulen, A., Jucker, F.M. & Pardi, A. Incorporating residual dipolar couplings into the NMR solution structure determination of nucleic acids. *Biopolymers* **52**, 168-180 (1999).

29. Hansen, M.R., Hanson, P. & Pardi, A. Filamentous bacteriophage for aligning RNA, DNA, and proteins for measurement of nuclear magnetic resonance dipolar coupling interactions. *Methods Enzymol.* **317**, 220-240 (2000).

30. Goddard, T.G. & Kneller, D.G. SPARKY 3. (University of California, San Francisco).

31. Guntert, P., Mumenthaler, C. & Wüthrich, K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273**, 283-298 (1997).

32. Zweckstetter, M. & Bax, A. Prediction of sterically induced alignment in a dilute liquid crystalline phase: Aid to protein structure determination by NMR. *J. Am. Chem. Soc.* **122**, 3791-3792 (2000).

33. Brünger, A.T. et al. Crystallography & NMR System: a new software suite for macromolecular structure determination. *Acta Cryst.* **D54**, 905-921 (1998).

34. Warren, J.J. & Moore, P.B. A maximum likelihood method for determining DaPQ and R for sets of dipolar coupling data. *J. Magn. Res.* **149**, 271-275 (2001).

35. Koradi, R., Billeter, M. & Wüthrich, K. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics* **14**, 51-55, 29-32 (1996).

36. Al-Hashimi, H.M., Gorin, A., Majumdar, A., Gosser, Y. & Patel, D.J. Towards structural genomics of RNA: rapid NMR resonance assignment and simultaneous RNA tertiary structure determination using residual dipolar couplings. *J. Mol. Biol.* **318**, 637-649 (2002).

37. Nicholls, A., Sharp, K.A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281-96. (1991).

38. Duarte, C.M. & Pyle, A.M. Stepping through an RNA structure: a novel approach to conformational analysis. *J. Mol. Biol.* **284**, 1465-1478 (1998).

Figure legends

Figure 1. Secondary structure and NMR data for domain 5 (D5). **A** Schematic representation of a group II intron with its six domains arranged around a central wheel. D5 is shown in red and the 5' and 3' exons are indicated in grey. The wild type sequence and secondary structure of D5 in the ai5γ group II intron from *S. cerevisiae* as used in this study. The different regions are colored as follows: The terminal helix 1 is blue; bulge nucleotides U9, A24, and C25 are green; G26 red; helix 2 and the GAAA tetraloop is shown in orange. Putative hydrogen bonding in the bulge region is indicated by dashed lines. 2'-endo sugar puckers (*) and nucleotides in a *syn*-conformation (**) are also indicated. **B** 600 MHz ^{1}H,^{1}H-NOESY spectrum of D5. The sequential H1'-H8/H6 assignments are traced with colored lines. The sequential walk is interrupted between nucleotides G26 and U27. The spectrum was acquired in D_2O at pH 6.5, 30 ºC, 0.5 mM RNA and 100 mM KCl. The NOESY mixing time was 250 ms.

Figure 2. Solution structure of D5. **A** View of the lowest energy structure. Helix 1 (nucleotides G1-A8, U27-C34) is colored in blue, helix 2 in orange (G10-C23), and the bulge nucleotides in green (U9, A24, C25) and red (*syn*-G26). **B** Calculated electrostatic surface potential of D5. Red indicates negative, white neutral and blue positive charge (red, –42; white –5; and blue, 5). In the panel to the right an unusual high electronegativity can be seen in the minor groove of the bulge region (-60 kT/e; arrow). Such a negative potential in the minor groove of the bulge is observed in all lowest energy structures (kT/e between –45 and –60). Range of negative potentials for other regions in D5 in the ten lowest energy structures are as follows: Major groove of bulge, –55 to –70 kT/e; major groove below bulge, –30 to –40 kT/e; major groove of tandem GC pairs, –30 to –45 kT/e; major groove of G3-U32 wobble, -70 to –80 kT/e; tetraloop, -35 to –45 kT/e.

Figure 3. Close up view of the bulge region. **A** Superposition of the bulge nucleotides (U9, A24-G26) as well as the neighboring base pairs (A8/U27, and G10/C23) of the ten lowest energy structures. A24 intercalates between U9 and G10, causing a tilt of the two nucleotides on the opposite strand. **B** A putative UC base pair in the bulge region is shown. The repulsion between the U9O2 and C25O2 could easily be overcome by coordination to either K^+ or Mg^{2+} ions. The exposure of the hydrophobic surface of the A8/U27 base pair and the strong "S"-shape of the backbone at the *syn*-G26 is visualized.

Figure 4. Change of chemical shifts in D5 upon Mg^{2+} binding. **A** Chemical shift changes are given for all aromatic protons in D5 upon addition of 7.5 mM $MgCl_2$. Bulge protons are affected most heavily. Coloring of the boxes corresponds to the secondary structure in Fig. 1A. **B** Chemical shift changes are given for H1' and H2' in D5 upon addition of 7.5 mM $MgCl_2$. Resonances in the bulge (A24 and C25) and at the opposite strand (U9) are affected most heavily. Coloring of the boxes corresponds to the secondary structure in Fig. 1A.

Legends of Supplementary Figures

Supplementary Figure 1. Superposition of all heavy atoms in D5 (A), in helix 1 (B) or helix 2 (C) of the ten lowest energy structures.

Supplementary Figure 2. Binding curve for Mg^{2+} as observed at H5 of C25. The experimental data is fit to a 1:1 binding curve, resulting in an individual binding constant $K_D^* = 3.3 \pm 0.7$ mM.

Supplementary Figure 3. Part of the $^1H,^1H$-NOESY spectrum of D5 in D_2O in the absence (A) and presence (B) of $MnCl_2$. In the presence of 20 μM MnCl2, the resonances involving A2 of the catalytic AGC triad are substantially broadend, whereas resonances of the neighboring G3 as well as U33 and C34 on the opposite strand are almost unaffected.

Table 1. Structural Statistics for the D5 structure[a]

	With dipolar coupling data	Without dipolar coupling data
NOE-derived distance restraints	678	678
Intranucleotide	283	283
Internucleotide ($\|i - j\| = 1$)	286	286
Long-range ($\|i - n\| \geq 2$)	104	104
repulsive	3	3
Dihedral restraints	261	261
Hydrogen bond restraints	72	72
Dipolar coupling restraints	24	0
r.m.s.d. (for all heavy atoms to mean structure(Å)		
Overall	4.15 ± 1.38	5.65 ± 1.69
Helix 1 (1-8, 27-34)	2.09 ± 0.67	2.25 ± 0.84
Bulge (9, 24-26)	1.70 ± 0.64	1.38 ± 0.37
Helix 2 & GAAA (10-23)	1.06 ± 0.25	1.06 ± 0.25
NOE Violations > 0.2 Å	0	0
Dihedral Violations > 5°	0	0
Average NOE r.m.s. deviation (Å)	0.021	0.014
Average Dihedral r.m.s. deviation (°)	0.932	0.635
Average RDC r.m.s. deviation (Hz)	2.07	--------

[a] Structural statistics are given for the 10 lowest energy structures out of 100 calculated structures.